

UNIT VI: Storage and Expression of Genetic Information

DNA Structure and Replication

29

I. OVERVIEW

Nucleic acids are required for the storage and expression of genetic information. There are two chemically distinct types of nucleic acids: deoxyribonucleic acid (DNA) and ribonucleic acid (RNA, see Chapter 30). DNA, the storehouse of genetic information, is present not only in chromosomes in the nucleus of eukaryotic organisms, but also in mitochondria and the chloroplasts of plants. Prokaryotic cells, which lack nuclei, have a single chromosome, but may also contain **nonchromosomal** DNA in the form of plasmids. The genetic information found in DNA is copied and transmitted to daughter cells through DNA replication. The DNA contained in a fertilized egg encodes the information that directs the development of an organism. This development may involve the production of billions of cells. Each cell is specialized, expressing only those functions that are required for it to perform its role in maintaining the organism. Therefore, DNA must be able to not only replicate precisely each time a cell divides, but also to have the information that it contains be selectively expressed. Transcription (RNA synthesis) is the first stage in the expression of genetic information (see Chapter 30). Next, the code contained in the nucleotide sequence of messenger RNA molecules is translated (protein synthesis, see Chapter 31), thus completing gene expression. This flow of information from DNA to RNA to protein is termed the "central dogma of molecular biology" (Figure 29.1), and is descriptive of all organisms, with the exception of some viruses that have RNA as the repository of their genetic information.

II. STRUCTURE OF DNA

DNA is a polydeoxyribonucleotide that contains many **monodeoxyribonucleotides** covalently linked by **3'→5'-phosphodiester bonds**. With the exception of a few viruses that contain single-stranded DNA, DNA exists as a **double-stranded molecule**, in which the two strands wind around each other, forming a **double helix**. In eukaryotic cells, DNA is found associated with various types of proteins (known collectively as **nucleoprotein**) present in the nucleus, whereas in prokaryotes, the protein-DNA complex is present in the **nucleoid**.

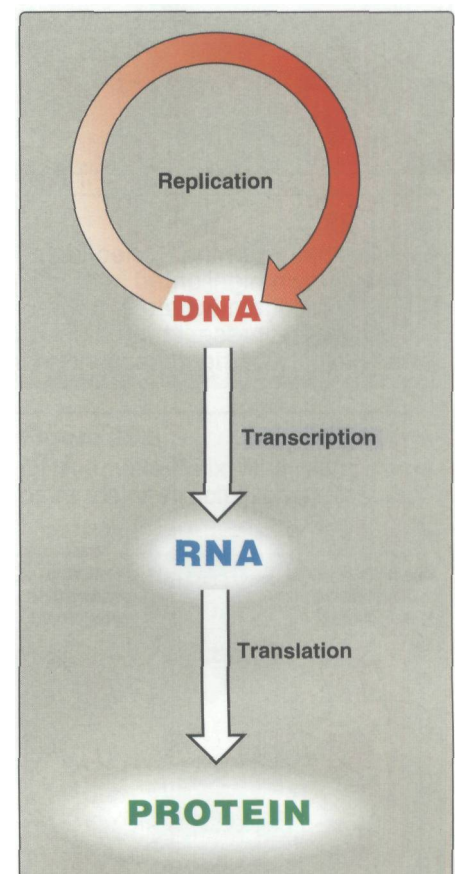
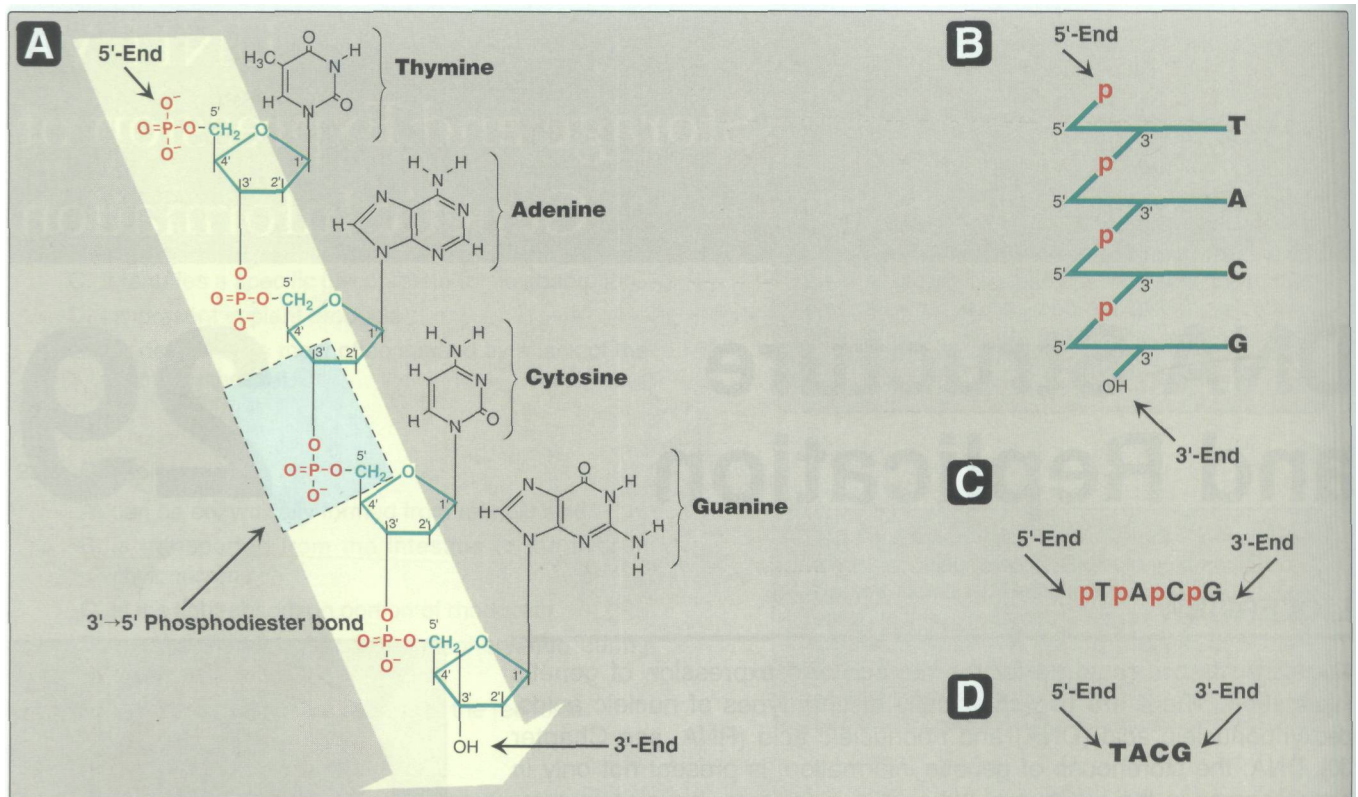


Figure 29.1
The "central dogma" of molecular biology.

**Figure 29.2**

A. DNA chain with the nucleotide sequence shown written in the 5' → 3' direction. A 3' → 5' phosphodiester bond is shown highlighted in the blue box, and the deoxyribose-phosphate backbone is shaded in yellow. B. The DNA chain written in a more stylized form, emphasizing the ribose-phosphate backbone. C. A simpler representation of the nucleotide sequence. D. The simplest representation, with the abbreviations for the bases written in the conventional 5' → 3' direction.

A. 3' → 5'-Phosphodiester bonds

Phosphodiester bonds join the 5'-hydroxyl group of the deoxypentose of one nucleotide to the 3'-hydroxyl group of the deoxypentose of an adjacent nucleotide through a phosphate group (Figure 29.2). The resulting long, unbranched chain has **polarity**, with both a 5'-end (the end with the free phosphate) and a 3'-end (the end with the free hydroxyl) that are not attached to other nucleotides. The bases located along the resulting deoxyribose-phosphate backbone are, by convention, always written in sequence from the 5'-end of the chain to the 3'-end. For example, the sequence of bases in the DNA shown in Figure 29.2 is read "thymine, adenine, cytosine, guanine" (5'-TACG-3'). Phosphodiester linkages between nucleotides (in DNA or RNA) can be cleaved hydrolytically by chemicals, or hydrolyzed **enzymatically** by a family of nucleases: **deoxyribonucleases** for DNA and **ribonucleases** for RNA. [Note: Nucleases that cleave the nucleotide chain at positions in the interior of the chain are called **endonucleases**. Those that cleave the chain only by removing individual nucleotides from one of the two ends are called **exonucleases**.]

B. Double helix

In the double helix, the two chains are coiled around a common axis called the **axis of symmetry**. The chains are paired in an **anti-parallel** manner, that is, the 5'-end of one strand is paired with the 3'-end of the other strand (Figure 29.3). In the DNA helix, the hydrophilic deoxyribose-phosphate backbone of each chain is on the outside of the molecule, whereas the hydrophobic bases are stacked inside. The overall structure resembles a twisted ladder. The spatial relationship between the two strands in the helix creates a **major (wide) groove** and a **minor (narrow) groove**. These grooves provide access for the binding of regulatory proteins to their specific recognition sequences along the DNA chain. [Note: Certain anticancer drugs, such as **dactinomycin (actinomycin D)**, exert their cytotoxic effect by intercalating into the narrow groove of the DNA double helix, thus interfering with RNA and DNA synthesis.¹]

1. Base pairing: The bases of one strand of DNA are paired with the bases of the second strand, so that an adenine is always paired with a thymine and a cytosine is always paired with a guanine. [Note: The base pairs are perpendicular to the axis of the helix (see Figure 29.3).] Therefore, one polynucleotide chain of the DNA double helix is always the **complement** of the other. Given the sequence of bases on one chain, the sequence of bases on the complementary chain can be determined (Figure 29.4). [Note: The specific base pairing in DNA leads to **Chargaff's Rules**: In any sample of double-stranded DNA, the amount of adenine equals the amount of thymine, the amount of guanine equals the amount of cytosine, and the total amount of purines equals the total amount of pyrimidines.] The base pairs are held together by **hydrogen bonds**: two between A and T and three between G and C (Figure 29.5). These hydrogen bonds, plus the hydrophobic interactions between the stacked bases, stabilize the structure of the double helix.

2. Separation of the two DNA strands in the double helix: The two strands of the double helix separate when hydrogen bonds between the paired bases are disrupted. Disruption can occur in the laboratory if the **pH** of the DNA solution is altered so that the nucleotide bases ionize, or if the solution is heated. [Note: Phosphodiester bonds are not broken by such treatment.] When DNA is heated, the temperature at which one half of the helical structure is lost is defined as the **melting temperature (T_m)**. The loss of helical structure in DNA, called **denaturation**, can be monitored by measuring its absorbance at 260 nm. [Note: Single-stranded DNA has a higher relative absorbance at this wavelength than does double-stranded DNA.] Because there are three hydrogen bonds between G and C but only two between A and T, DNA that contains high concentrations of A and T denatures at a lower temperature than G- and C-rich DNA (Figure 29.6). Under appropriate conditions, complementary DNA strands can reform the double helix by the process called **renaturation** (or **reannealing**).

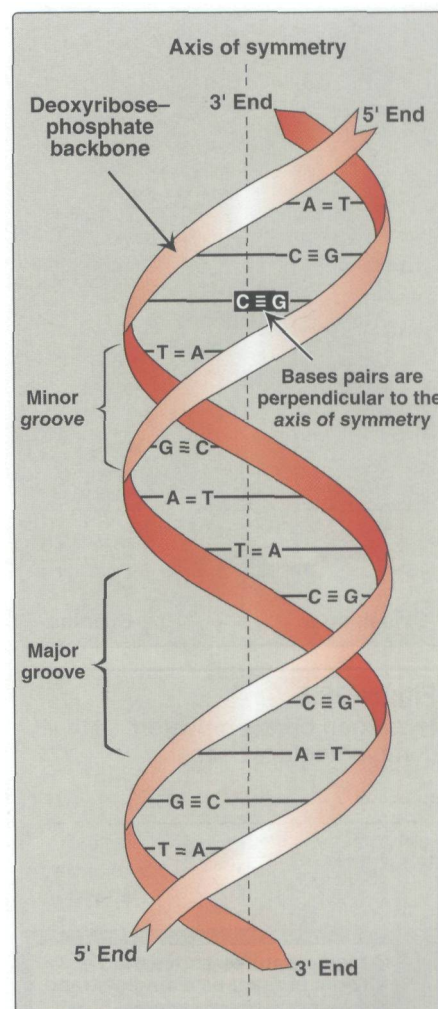


Figure 29.3

DNA double helix, illustrating some of its major structural features.

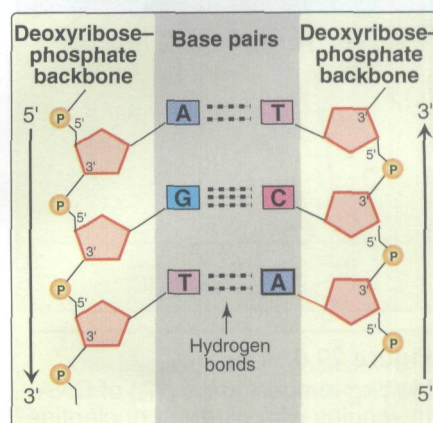


Figure 29.4

Two complementary DNA sequences.



¹See Chapter 40 in *Lippincott's Illustrated Reviews: Pharmacology* (3rd Ed.) and Chapter 38 in (2nd Ed.) for a discussion of the anticancer drug, actinomycin D.

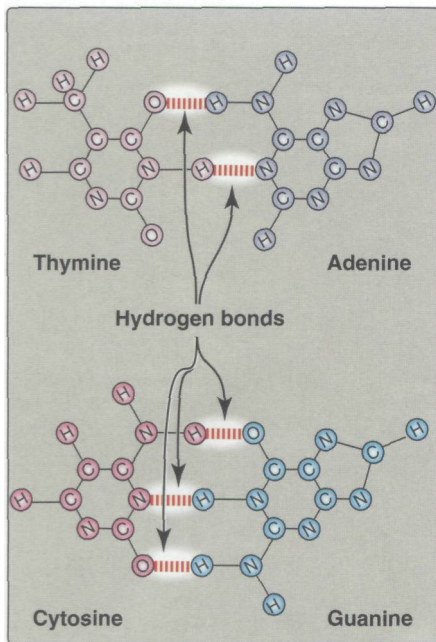


Figure 29.5
Hydrogen bonds between complementary bases.

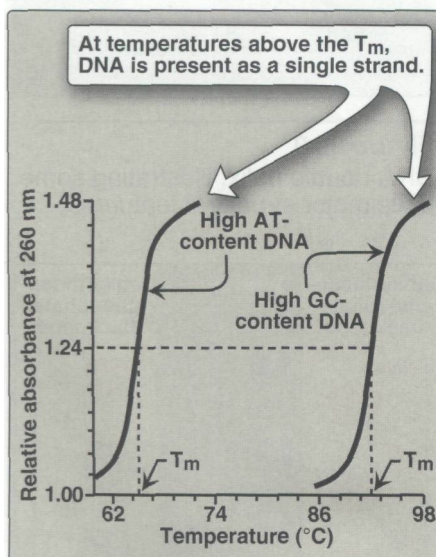


Figure 29.6
Melting temperatures (T_m) of DNA molecules with different nucleotide compositions. (At a wavelength of 260 nm, single-stranded DNA has a higher relative absorbance than does double-stranded DNA.)

3. Structural forms of the double helix: There are three major structural forms of DNA: the B form, described by Watson and Crick in 1953, the A form, and the Z form. The B form is a right-handed helix with ten residues per 360° turn of the helix, and with the planes of the bases perpendicular to the helical axis. Chromosomal DNA is thought to consist primarily of **B-DNA** (Figure 29.7 illustrates a space-filling model of B-DNA). The A form is produced by moderately dehydrating the B form. It is also a right-handed helix, but there are eleven base pairs per turn, and the planes of the base pairs are tilted 20° away from the perpendicular to the helical axis. The conformation found in DNA-RNA hybrids or RNA-RNA double-stranded regions is probably very close to the A form. **Z-DNA** is a left-handed helix that contains about twelve base pairs per turn (see Figure 29.7). [Note: The deoxyribose-phosphate backbone "zigzags," hence, the name "Z"-DNA.] Stretches of Z-DNA can occur naturally in regions of DNA that have a sequence of alternating purines and pyrimidines, for example, poly GC. Transitions between the helical forms of DNA may play an important role in regulating gene expression.

C. Circular DNA molecules

Each chromosome in the nucleus of a eukaryote contains one long linear molecule of double-stranded DNA, which is bound to a complex mixture of proteins to form **chromatin**. Eukaryotes have closed circular DNA molecules in their mitochondria, as do plant **chloroplasts**. A prokaryotic organism contains a single, double-stranded, supercoiled, circular chromosome. Each prokaryotic chromosome is associated with **histone-like** proteins (see p. 406) and RNA that can condense the DNA to form a **nucleoid**. In addition, most species of bacteria also contain small, circular, **extrachromosomal** DNA molecules called **plasmids**. Plasmid DNA carries genetic information, and undergoes replication that may or may not be synchronized to chromosomal division.² Plasmids may carry genes that convey antibiotic resistance to the host bacterium, and may facilitate the transfer of genetic information from one bacterium to another. [Note: The use of plasmids as vectors in recombinant DNA technology is described in Chapter 32.]

III. STEPS IN PROKARYOTIC DNA SYNTHESIS

When the two strands of the DNA double helix are separated, each can serve as a template for the replication of a new complementary strand. This produces two daughter molecules, each of which contains two DNA strands with an antiparallel orientation (see Figure 29.3). This process is called **semiconservative replication** because, although the parental duplex is separated into two halves (and, therefore, is not "conserved" as an entity), each of the individual parental strands remains intact in **one** of the two new duplexes (Figure 29.8). The enzymes involved in the DNA replication process are template-directed polymerases that can synthesize the complementary sequence of each strand with extraordinary fidelity. The reactions described in this section were first known from



²See p.116 in *Lippincott's Illustrated Reviews: Microbiology* for a discussion of plasmids.

studies of the bacterium *Escherichia coli* (*E. coli*), and the description given below refers to the process in that microorganism. DNA synthesis in higher organisms is less well understood, but involves the same types of mechanisms. In either case, initiation of DNA replication commits the cell to continue the process until the entire genome has been replicated.

A. Separation of the two complementary DNA strands

In order for the two strands of the parental double helical DNA to be replicated, they must first separate (or "melt"), at least in a small region, because the polymerases use only single-stranded DNA as a template. In prokaryotic organisms, DNA replication begins at a single, unique nucleotide **sequence**—a site called the **origin of replication** (Figure 29.9 A). In eukaryotes, replication begins at multiple sites along the DNA helix (Figure 29.9 B). These sites include a short sequence composed almost exclusively of AT base pairs. [Note: This is referred to as a **consensus sequence**, because the order of nucleotides is essentially the same at each site.] Having multiple origins of replication provides a mechanism for rapidly replicating the great length of the eukaryotic DNA molecules.

B. Formation of the replication fork

As the two strands unwind and separate they form a "V" where active synthesis occurs. This region is called the **replication fork**. It moves along the DNA molecule as synthesis occurs. Replication of double-stranded DNA is **bidirectional**—that is, the replication forks move in both directions away from the origin (see Figure 29.9).

1. Proteins required for DNA strand separation: Initiation of DNA replication requires the recognition of the origin of replication and/or the replication fork by a group of proteins that form the **prepriming complex**. These proteins are responsible for maintaining the separation of the parental strands, and for unwinding the double helix ahead of the advancing replication fork. These proteins include the following:

a. DnaA protein: Twenty to fifty monomers of dnaA protein bind to specific nucleotide sequences at the origin of replication, which is particularly rich in AT base pairs. This ATP-requiring process causes the double-stranded DNA to **melt**—that is, the strands separate, forming localized regions of single-stranded DNA.

b. Single-stranded DNA-binding (SSB) proteins: Also called **helix-destabilizing proteins**, these bind only to single-stranded DNA (Figure 29.10). They bind **cooperatively**—that is, the binding of one molecule of SSB protein makes it easier for additional molecules of SSB protein to bind tightly to the DNA strand. The SSB proteins are not enzymes, but rather serve to shift the equilibrium between double- and single-stranded DNA in the direction of the single-stranded forms. These proteins not only keep the two strands of DNA separated in the area of the replication origin, thus providing the single-stranded template required by polymerases, **but also** protect the DNA from nucleases that cleave single-stranded DNA.

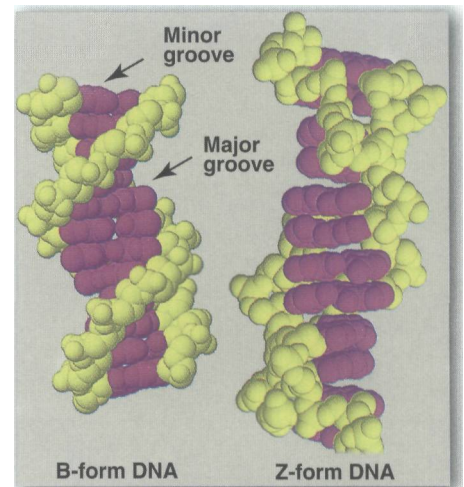


Figure 29.7
Structures of B-DNA and Z-DNA.

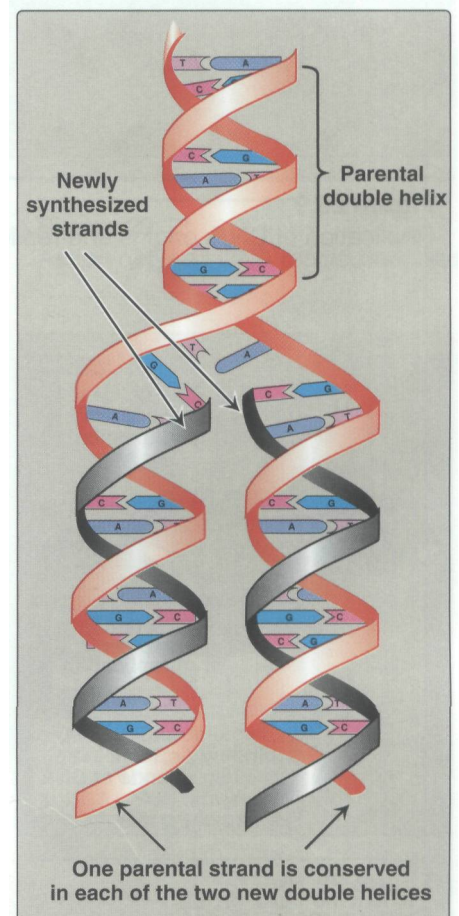
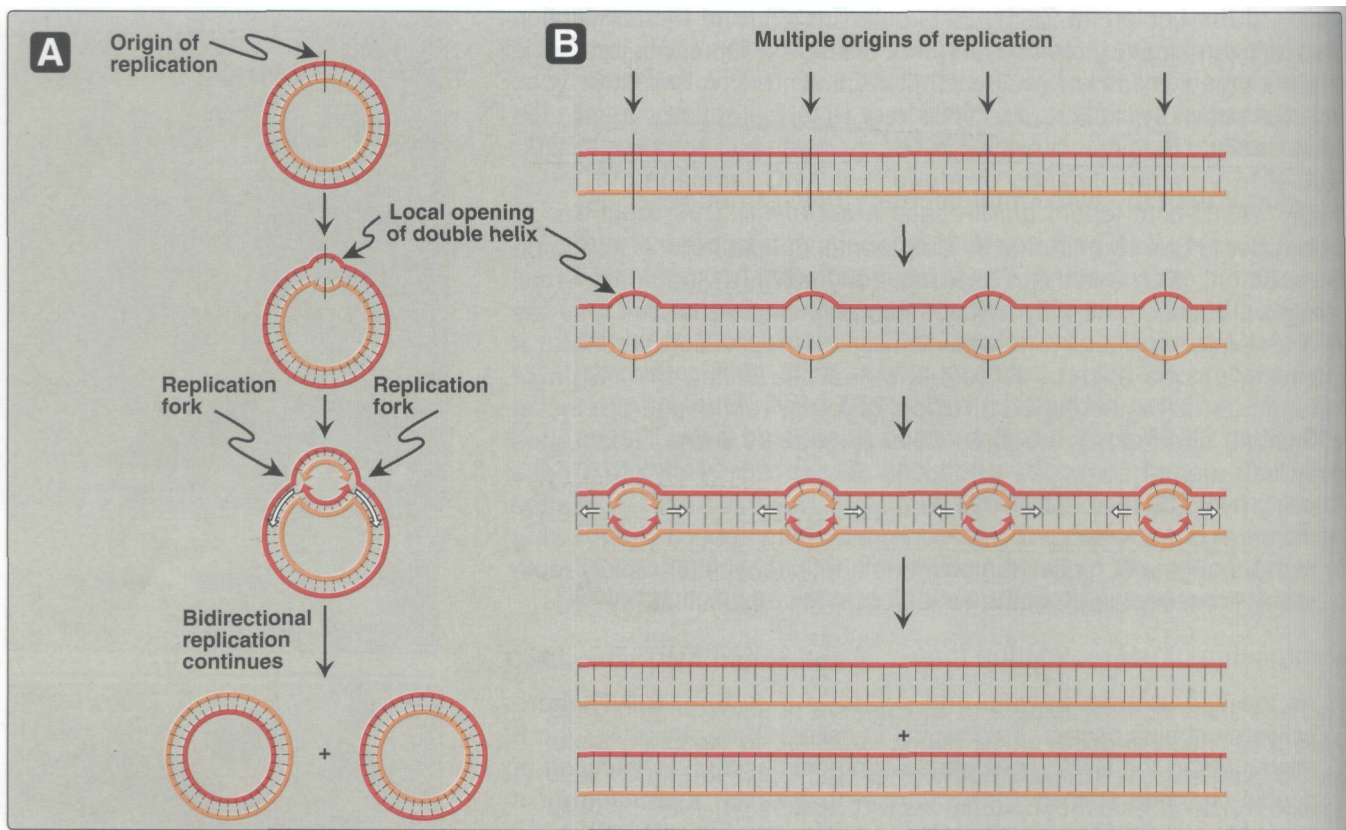
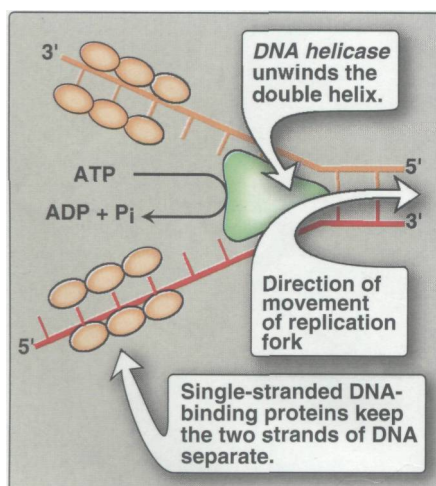


Figure 29.8
Semiconservative replication of DNA.

**Figure 29.9**

Replication of DNA: origins and replication forks. A. Small prokaryotic circular DNA. B. Very long eukaryotic DNA.

**Figure 29.10**

Proteins responsible for maintaining the separation of the parental strands and unwinding the double helix ahead of the advancing replication fork.

c. DNA helicases: These enzymes bind to single-stranded DNA near the replication fork, and then move into the neighboring double-stranded region, forcing the strands **apart**—in effect, unwinding the double helix. *Helicases* require energy provided by ATP. When the strands separate, SSB proteins bind, preventing reformation of the double helix (see Figure 29.10).

2. Solving the problem of supercoils: As the two strands of the double helix are separated, a problem is encountered, namely, the appearance of **positive supercoils** (also called **supertwists**) in the region of DNA ahead of the replication fork (Figure 29.11). The accumulating positive supercoils interfere with further unwinding of the double helix. [Note: Supercoiling can be demonstrated by tightly grasping one end of a telephone cord while twisting the other end. If the cord is twisted in the direction of tightening the coils, the cord will wrap around itself in space to form positive supercoils. If the cord is twisted in the direction of loosening the coils, the cord will wrap around itself in the opposite direction to form **negative supercoils**.] To solve this problem, there is a group of enzymes called **DNA topoisomerases**, which are responsible for removing supercoils in the helix.

a. Type I DNA topoisomerases reversibly cut a single strand of the double helix. They have both **nuclease (strand-cutting)** and **ligase (strand-resealing)** activities. They do not require ATP, but rather appear to store the energy from the **phosphodiester**

ester bond they cleave, reusing the energy to reseal the strand (Figure 29.12). Each time a transient "nick" is created in one DNA strand, the intact DNA strand is passed through the break before it is resealed, thus relieving ("relaxing") accumulated supercoils. *Type I topoisomerases* relax negative supercoils (that is, those that contain fewer turns of the helix than relaxed DNA) in *E. coli*, and both negative and positive supercoils (that is, those that contain fewer or more turns of the helix than relaxed DNA) in eukaryotic cells.

- b. **Type II DNA topoisomerases** bind tightly to the DNA double helix and make transient breaks in both strands. The enzyme then causes a second stretch of the DNA double helix to pass through the break and, finally, reseals the break (Figure 29.13). As a result, both negative and positive supercoils can be relieved. *Type II DNA topoisomerases* are also required in both prokaryotes and eukaryotes for the separation of interlocked molecules of DNA following chromosomal replication. [Note: **Anticancer agents**, such as **etoposide**,³ target human *topoisomerase II*.] *DNA gyrase*, a *type II topoisomerase* found in *E. coli*, has the unusual property of being able to introduce negative supercoils into relaxed circular DNA using energy from the hydrolysis of ATP. This facilitates the future replication of DNA because the negative supercoils neutralize the positive supercoils introduced during opening of the double helix. It also aids in the transient strand separation required during transcription (see p. 416). [Note: Bacterial *DNA gyrase* is the unique target of a group of **antimicrobial agents** called quinolones, for example, **ciprofloxacin**.⁴]

C. Direction of DNA replication

The DNA polymerases responsible for copying the DNA templates are only able to "read" the parental nucleotide sequences in the 3'→5' direction, and they synthesize the new DNA strands in the 5'→3' (antiparallel) direction. Therefore, beginning with one parental double helix, the two newly synthesized stretches of nucleotide chains must grow in opposite directions—one in the 5'→3' direction toward the replication fork and one in the 5'→3' direction away from the replication fork (Figure 29.14). This feat is accomplished by a slightly different mechanism on each strand.

1. **Leading strand:** The strand that is being copied in the direction of the advancing replication fork is called the leading strand and is **synthesized almost continuously**.
2. **Lagging strand:** The strand that is being copied in the direction away from the replication fork is **synthesized discontinuously**, with small fragments of DNA being copied near the replication fork. These short stretches of discontinuous DNA, termed **Okazaki fragments**, are eventually joined to become a single, continuous strand. The new strand of DNA produced by this mechanism is termed the lagging strand.

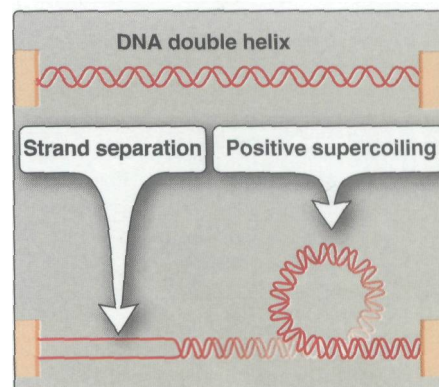


Figure 29.11
Positive supercoiling resulting from DNA strand separation.

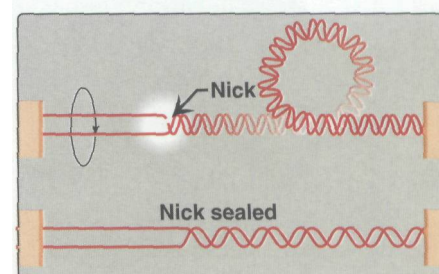


Figure 29.12
Action of type I DNA topoisomerases

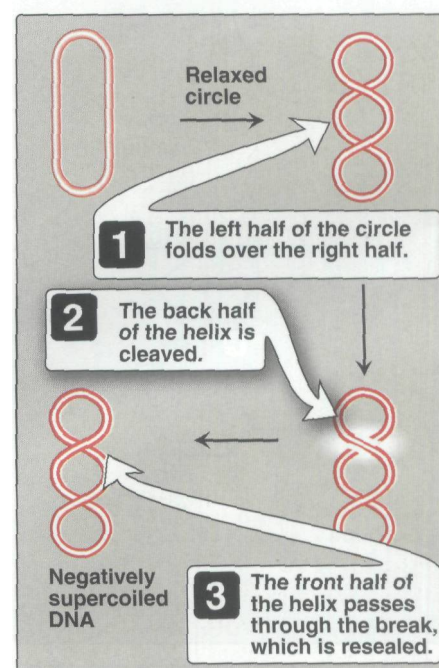


Figure 29.13
Action of type II DNA topoisomerase.



³See Chapter 40 in *Lippincott's Illustrated Reviews: Pharmacology* (3rd Ed.) and Chapter 38 (2nd Ed.) for a discussion of etoposide as an anticancer agent.

⁴See Chapter 34 in *Lippincott's Illustrated Reviews: Pharmacology* (3rd Ed.) and Chapter 32 (2nd Ed.) for a discussion of the quinolones.

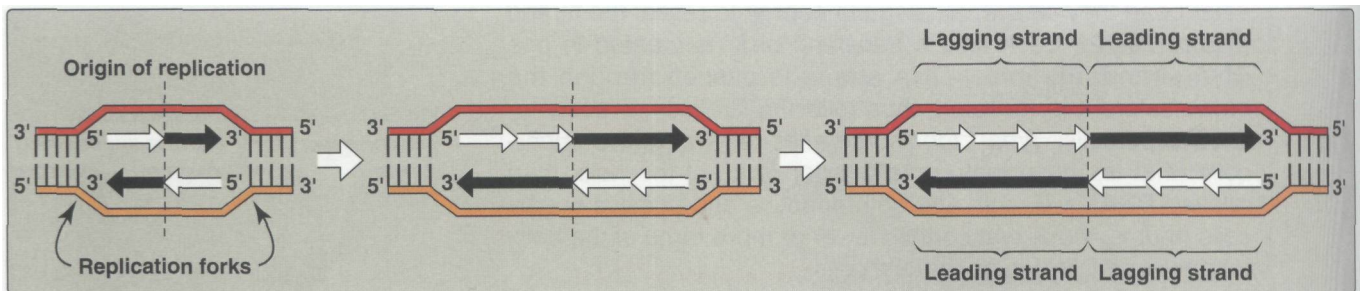


Figure 29.14
Discontinuous synthesis of DNA.

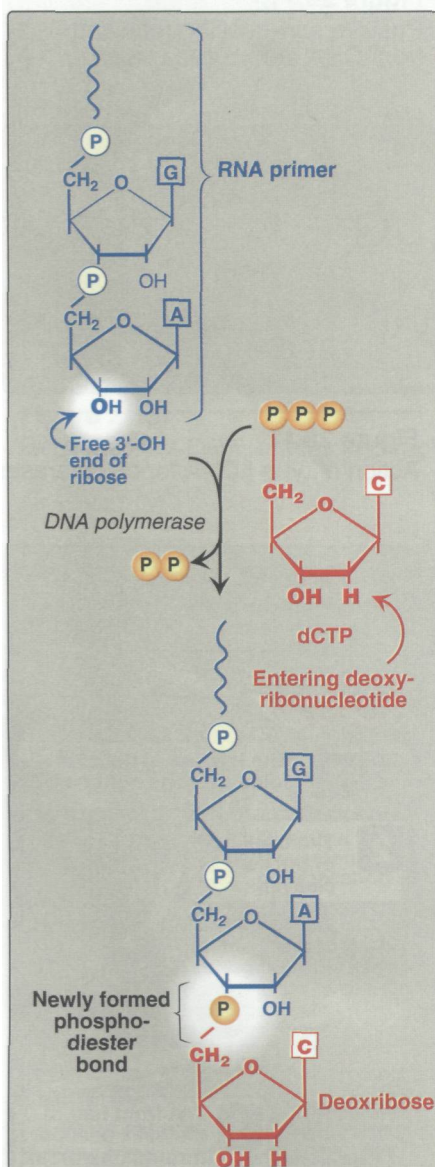


Figure 29.15
Use of an RNA primer to initiate DNA synthesis.

D. RNA primer

DNA polymerases cannot initiate synthesis of a complementary strand of DNA on a totally single-stranded template. Rather, they require an RNA primer—that is, a short, double-stranded region consisting of RNA base-paired to the DNA template, with a free hydroxyl group on the 3'-end of the RNA strand (Figure 29.15). This hydroxyl group serves as the first acceptor of a nucleotide by action of DNA polymerase. In *de novo* DNA synthesis, that free 3'-hydroxyl is provided by the short stretch of RNA, rather than DNA.

- 1. Primase:** A specific **RNA polymerase**, called *primase*, synthesizes the short stretches of RNA (approximately ten nucleotides long) that are complementary and antiparallel to the DNA template. In the resulting **hybrid duplex**, the U in RNA pairs with A in DNA. As shown in Figure 29.16, these short RNA sequences are constantly being synthesized at the replication fork on the lagging strand, but only one RNA sequence at the origin of replication is required on the leading strand. The building blocks for this process are **5'-ribonucleoside triphosphates**, and pyrophosphate is released as each phosphodiester bond is made. [Note: The RNA primer is later removed as described on p. 402.]
- 2. Primosome:** Prior to the beginning of RNA primer synthesis on the lagging strand, a prepriming complex of several proteins is assembled and binds to the single strand of DNA, displacing some of the single-stranded DNA-binding proteins. This protein complex, plus *primase*, is called the **primosome**. It initiates Okazaki fragment formation by moving along the template for the lagging strand in the 5'→3' direction, periodically recognizing specific sequences of nucleotides that direct it to create an RNA primer that is synthesized in the 5'→3' direction (antiparallel to the DNA template chain).

E. Chain elongation

Prokaryotic and eukaryotic DNA polymerases elongate a new DNA strand by adding deoxyribonucleotides, one at a time, to the 3'-end of the growing chain (see Figure 29.16). The sequence of nucleotides that are added is dictated by the base sequence of the template strand with which the incoming nucleotides are paired.

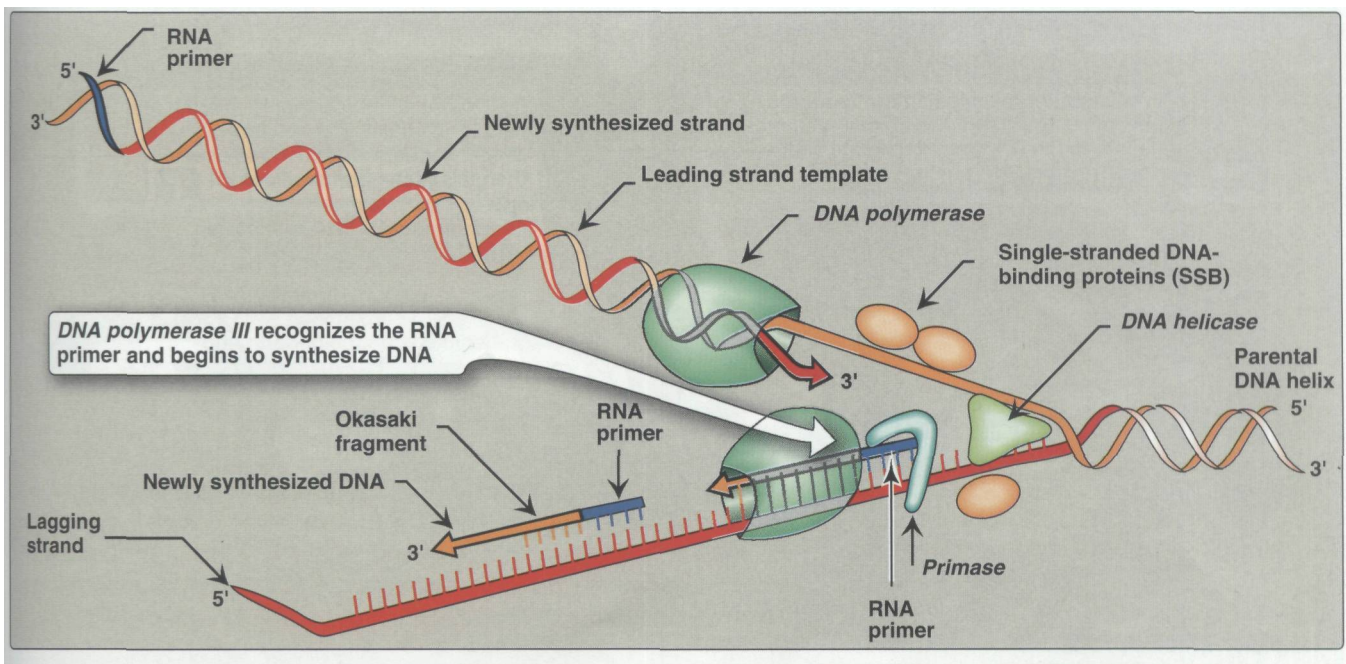


Figure 29.16
Elongation of the leading and lagging strands.

1. **DNA polymerase III:** DNA chain elongation is catalyzed by *DNA polymerase III*. Using the 3'-hydroxyl group of the RNA primer as the acceptor of the first deoxyribonucleotide, *DNA polymerase III* begins to add nucleotides along the single-stranded template that specifies the sequence of bases in the newly synthesized chain. *DNA polymerase III* is a highly "processive" enzyme—that is, it remains bound to the template strand as it moves along, and does not have to diffuse away and rebind before adding each new nucleotide. The new strand grows in the 5'→3' direction, antiparallel to the parental strand (see Figure 29.16). The nucleotide building blocks are **5'-deoxyribonucleoside triphosphates**. Pyrophosphate (PP_i) is released when each new nucleotide is added to the growing chain (see Figure 29.15). [Note: The further hydrolysis of pyrophosphate to two phosphates means that a total of two high-energy bonds are used to drive the addition of each deoxynucleotide.] All four deoxyribonucleoside triphosphates (dATP, dTTP, dCTP, and dGTP) must be present for DNA elongation to occur. If one of the four is in short supply, DNA synthesis stops when that nucleotide is depleted.
2. **Proofreading of newly synthesized DNA:** It is highly important for the survival of an organism that the nucleotide sequence of DNA be replicated with as few errors as possible. Misreading of the template sequence could result in deleterious, perhaps lethal, mutations. To ensure replication fidelity, *DNA polymerase III* has, in addition to its 5'→3' polymerase activity, a "proofreading" activity (3'→5' exonuclease, Figure 29.17). As each nucleotide is added to the chain, *DNA polymerase III* checks to make certain the added

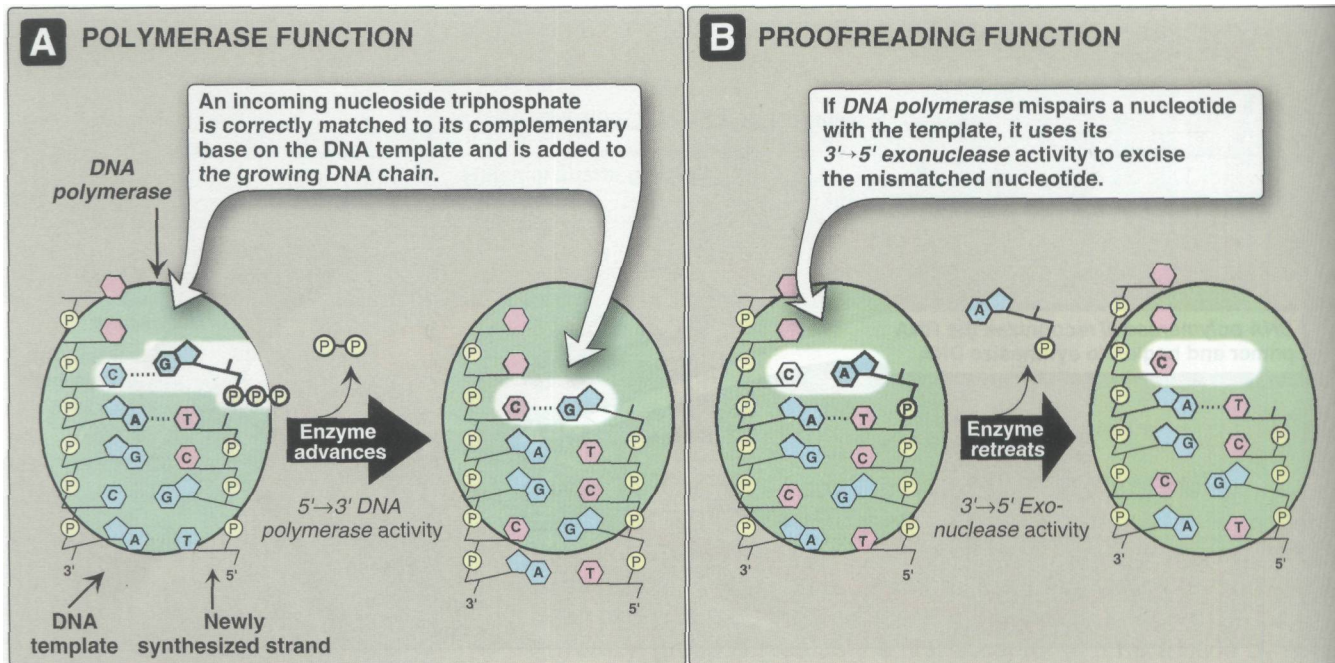


Figure 29.17

3'→5'-Exonuclease activity enables DNA polymerase III to "proofread" the newly synthesized DNA strand.

nucleotide is, in fact, correctly matched to its complementary base on the template. If it is not, the 3'→5' exonuclease activity edits the mistake. [Note: The enzyme requires an improperly base-paired 3'-hydroxy terminus and, therefore, does not degrade correctly paired nucleotide sequences.] For example, if the template base is cytosine and the enzyme mistakenly inserts an adenine instead of a guanine into the new chain, the 3'→5' exonuclease hydrolytically removes the misplaced nucleotide. The 5'→3' polymerase then replaces it with the correct nucleotide containing guanine (see Figure 29.17). [Note: The proofreading activity requires an exonuclease that moves in the 3'→5' direction, not 5'→3' like the polymerase activity. This is because the excision must be done in the reverse direction from that of synthesis.]

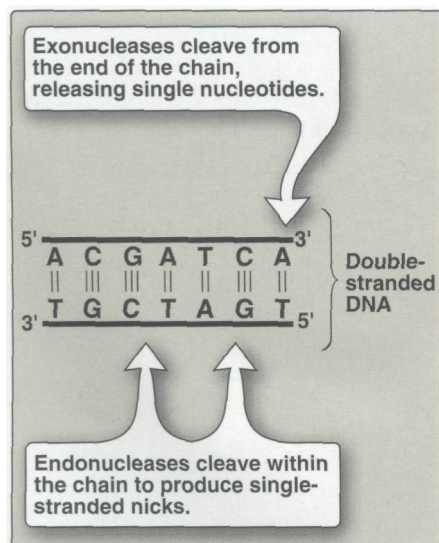


Figure 29.18

Endonuclease versus exonuclease activity.

F. Excision of RNA primers and their replacement by DNA

DNA polymerase III continues to synthesize DNA on the lagging strand until it is blocked by proximity to an RNA primer. When this occurs, the RNA is excised and the gap filled by DNA polymerase I.

- 5'→3' Exonuclease activity:** In addition to having the 5'→3' polymerase activity that synthesizes DNA, and the 3'→5' exonuclease activity that proofreads the newly synthesized DNA chain like DNA polymerase III, DNA polymerase I also has a 5'→3' exonuclease activity that is able to hydrolytically remove the RNA primer. [Note: These activities are **exonucleases** because they remove one nucleotide at a time from the end of the DNA chain rather than cleaving it internally as do the **endonucleases** (Figure 29.18).] First, DNA polymerase I locates the space ("nick" between the 3'-end of the DNA newly synthesized by DNA polymerase III and the 5'-end of the adjacent RNA primer. Next, DN

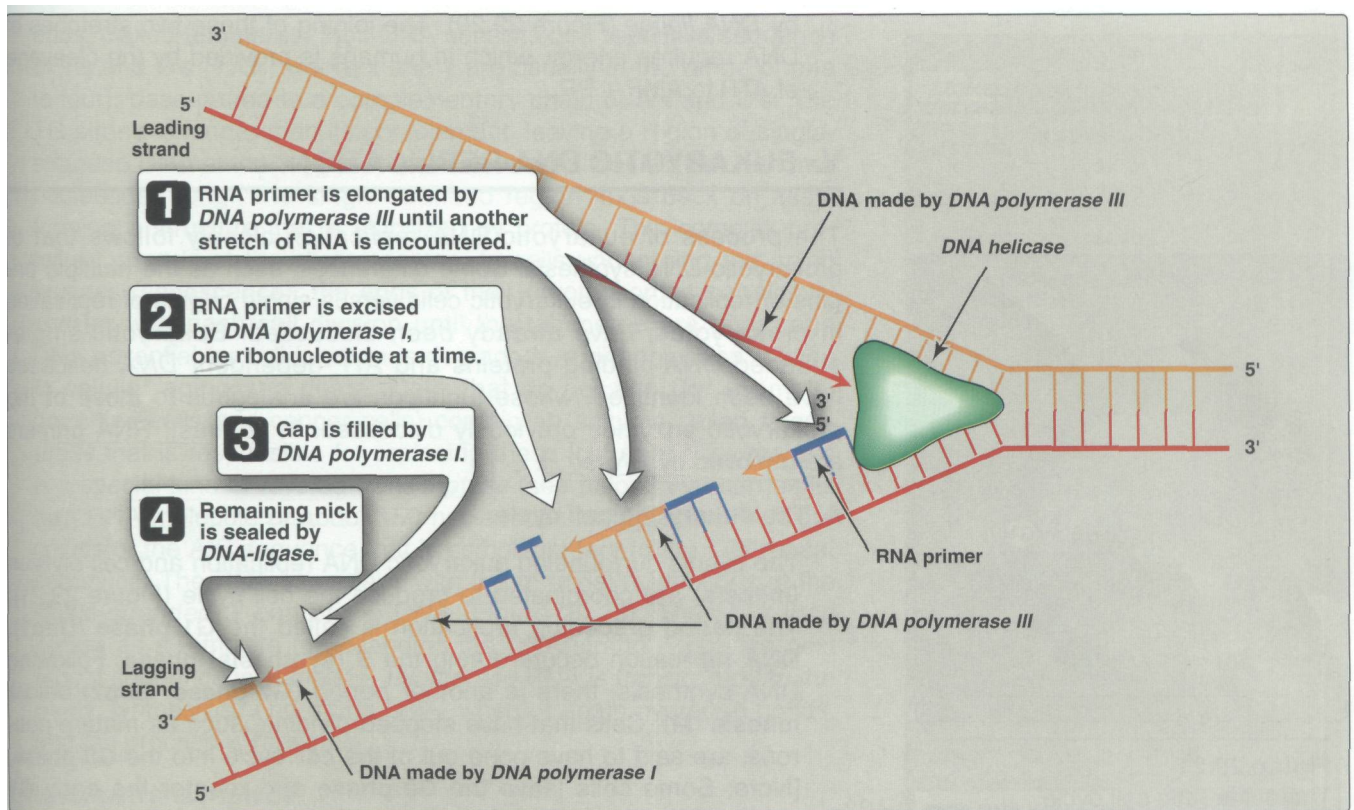


Figure 29.19

Removal of RNA primer and filling of the resulting "gaps" by DNA polymerase I.

polymerase I hydrolytically removes the RNA nucleotides "ahead" of itself, moving in the $5' \rightarrow 3'$ direction ($5' \rightarrow 3'$ **exonuclease activity**). As it removes the RNA, *DNA polymerase I* replaces it with deoxyribonucleotides, synthesizing DNA in the $5' \rightarrow 3'$ direction ($5' \rightarrow 3'$ **polymerase activity**). As it synthesizes the DNA, it also "proofreads" the new chain using $3' \rightarrow 5'$ **exonuclease activity**. This removal/synthesis/proofreading continues, one nucleotide at a time, until the RNA is totally degraded and the gap is filled with DNA (Figure 29.19).

Differences between $5' \rightarrow 3'$ and $3' \rightarrow 5'$ exonucleases: The $5' \rightarrow 3'$ exonuclease activity of *DNA polymerase I* differs from the $3' \rightarrow 5'$ exonuclease used by both *DNA polymerase I* and *III* in two important ways. First, $5' \rightarrow 3'$ exonuclease can remove one nucleotide at a time from a region of DNA that is properly base-paired. The nucleotides it removes can be either ribonucleotides or deoxyribonucleotides. Second, $5' \rightarrow 3'$ exonuclease can also remove groups of altered nucleotides in the $5' \rightarrow 3'$ direction, removing from one to ten nucleotides at a time. This ability is important in the repair of some types of damaged DNA.

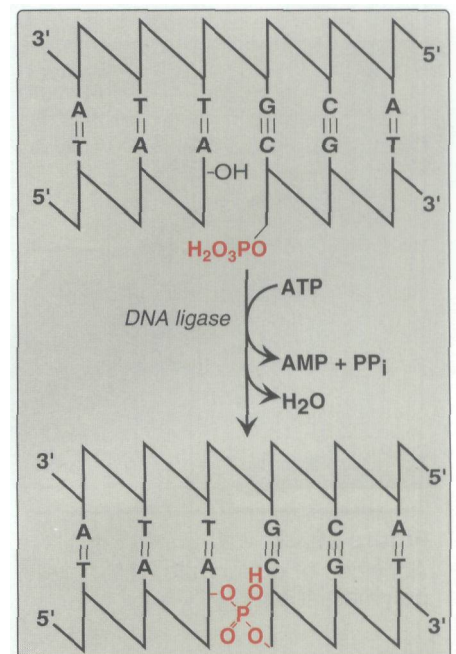


Figure 29.20

Formation of a phosphodiester bond by DNA ligase.

G. DNA ligase

The final phosphodiester linkage between the 5'-phosphate group on the DNA chain synthesized by *DNA polymerase III* and the 3'-hydroxyl group on the chain made by *DNA polymerase I* is catalyzed